

Joint Source Estimation and Localization

Souleymen Sahnoun and Pierre Comon, *Fellow, IEEE*

Abstract

The estimation of directions of arrival is formulated as the decomposition of a 3-way array into a sum of rank-one terms, which is possible when the receive array enjoys some geometrical structure. The main advantage is that this decomposition is essentially unique under mild assumptions, if computed exactly. The drawback is that a low-rank approximation does not always exist. Therefore, a coherence constraint is introduced that ensures the existence of the latter best approximate, which allows to localize and estimate closely located or highly correlated sources. Then Cramér-Rao bounds are derived for localization parameters and source signals, assuming the others are nuisance parameters; some inaccuracies found in the literature are pointed out. Performances are eventually compared with unconstrained reference algorithms such as ESPRIT, in the presence of additive complex Gaussian noise, with possibly non circular distribution.

Index Terms

multi-way array, coherence, tensor decomposition, source localization, antenna array processing, low-rank approximation, complex Cramér-Rao bounds, non circularity

I. INTRODUCTION

Estimation of Directions of Arrival (DoA) is a central problem in antenna array processing, including in particular radar, sonar, or telecommunications [2]. Over the last decades, several DoA estimation tools have been developed, ranging from nonparametric Fourier-based methods to parametric high-resolution techniques. The latter techniques, including linear prediction-based methods and subspace methods, are often preferred to nonparametric ones since they achieve high resolution estimates. Recently, methods based on sparse approximations have been proposed, which are considered as semi-parametric [3], [4].

Traditional subspace approaches such as MUSIC (multiple signal classification) are based on low-rank approximation of the covariance matrix of observations, and on detecting points of minimal distance with the so-called array manifold [5] [6]. These approaches hence assume that (i) the measurements are weakly stationary over sufficiently long observation lengths, (ii) the number of sources of interest is smaller than the number of sensors, and (iii) spatial

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

This work has been funded by the FP7 programme 2007–2013, ERC Grant Agreement no. 320594, DECODA project. It has been partly presented during the 2014 SAM Workshop [1].

S. Sahnoun and P. Comon are with the Gipsa-Lab, 11 rue des Mathématiques, Domaine Universitaire BP 46, 38402 Saint Martin d'Hères Cedex, France (e-mail: firstname.name@gipsa-lab.grenoble-inp.fr)

responses of all sensors are known, as well as their location (in other words, the sensor array needs to be calibrated). Another well-known subspace method is the ESPRIT (estimation of signal parameters via rotational invariance techniques) algorithm [7] which is applicable to sensor arrays consisting of two identical displaced subarrays. The displacement vector between the two subarrays should be known, whereas, unlike MUSIC-type algorithms, the calibration information of each subarray is not required.

In [8], an extension of ESPRIT to more than two subarrays has been proposed, but translations were collinear. In [9], an extension of root-MUSIC has been proposed, and applied to parallel linear subarrays. In the present paper, as in [10], we consider an arbitrary number of subarrays, deduced from each other by arbitrary translations. In [10], a deterministic approach has been proposed, which permits not only to work with short data lengths (and hence less stationary sources), but also to localize more sources than sensors present in the reference array. This approach is based on the same rotational invariance as exploited in ESPRIT [7], but can handle more than one displacement. It consists in storing the measurements in a 3-way array, and to decompose it into a sum of rank-one terms. One very interesting by-product of [10] is that source estimates are also delivered for free, without any further estimation stage.

The most popular model for fitting a 3-way array using a sum of rank-one terms is the canonical polyadic (CP) decomposition. Contrary to the decomposition of matrices in rank-one terms, the CP decomposition is essentially unique under mild conditions. However, a best-fitting CP decomposition may not exist for some tensors [11]. In this case, trying to find a low-rank tensor approximation results in diverging components.

In this paper we revisit the approach of [10], where the important issue of the existence of a low-rank tensor approximation has been neglected, so that the latter approach is actually ill posed. This fact has been already pointed out in [11], and additional constraints have been suggested, which involve the coherences of factor matrices, and enjoy a reliable physical meaning and at the same time ensure existence of a solution. In the present work, we shed some light on conditioning and algorithmic issues. Our main contributions are the following.

We first propose a new differentiable coherence constraint which guarantees existence of the low-rank tensor approximation. The latter constraint is implemented in the form of penalty in a descent algorithm. Our approach is general and could be applied to other fields than antenna array processing, which is chosen as an illustration.

Then, we derive the expressions of the Cramér-Rao bounds (CRB) related to localization and estimation problems in the presence of nuisance parameters estimated by CP decomposition methods. Note that the CRB for CP decomposition of complex-valued and real-valued tensors were respectively studied in [12] and [13], but without assuming that one of matrix factors is parameterized by angles of arrival.

The remainder of this paper is organized as follows. In the next section, we introduce notation and formulate the DoA estimation as a low-rank decomposition of a 3-way array. Then we define in Section III the coherence constraint ensuring existence of the best approximation. An optimization algorithm is described in Section IV, and Cramér-Rao bounds of DoA parameters are derived in Section V. Finally, we show the benefit of the proposed algorithm via numerical experiments, and compare results to Cramér-Rao bounds.

II. MODELING AND NOTATION

In the following, vectors will be denoted by bold lowercases, *e.g.* \mathbf{a} , whereas matrices or higher-order arrays will be denoted by bold uppercases, *e.g.* \mathbf{A} . Moreover, \mathbf{a}_r will denote the r th column of matrix \mathbf{A} .

Suppose that R narrow-band radiating sources impinge on an array of sensors, formed of L subarrays of K sensors each, with arbitrary shape. We make the far-field assumption, that is, we assume that sources are located sufficiently far from the array, compared to the array dimensions, so that waves can be considered as plane. The key assumption made in [7], [10], [11] is that, taking one subarray as reference, every subarray can be deduced from the reference one by an unknown translation in space, defined by some vector $\boldsymbol{\delta}_\ell$ of \mathbb{R}^3 , $1 \leq \ell \leq L$, $\boldsymbol{\delta}_1 \stackrel{\text{def}}{=} \mathbf{0}$.

Denote $\varsigma_r(t)$ the signal transmitted by the r th source, ψ_r its DoA viewed from the array, \mathbf{d}_r its associated unit modulus directional vector in \mathbb{R}^3 , and $s(\mathbf{y}, t)$ the signal measured at a point in space defined by its coordinates \mathbf{y} (we consider complex envelopes about the central frequency). Then we have:

$$s(\mathbf{y}, t) = \sum_{r=1}^R \varsigma_r(t) a_r(\mathbf{y}), \quad a_r(\mathbf{y}) \stackrel{\text{def}}{=} \exp\{j \frac{\omega}{C} \mathbf{y}^\top \mathbf{d}_r\} \quad (1)$$

where ω is the central pulsation of the narrow-band waves, C the wave velocity, and $j = \sqrt{-1}$. Because waves are plane and narrow-band, the signal measured at another point $\mathbf{y} + \boldsymbol{\tau}$, deduced from \mathbf{y} by a translation $\boldsymbol{\tau}$ takes the form:

$$s(\mathbf{y}, \boldsymbol{\tau}, t) = \sum_{r=1}^R a_r(\mathbf{y}) b_r(\boldsymbol{\tau}) \varsigma_r(t), \quad b_r(\boldsymbol{\tau}) \stackrel{\text{def}}{=} \exp\{j \frac{\omega}{C} \boldsymbol{\tau}^\top \mathbf{d}_r\} \quad (2)$$

In other words, function $s(\mathbf{y}, \boldsymbol{\tau}, t)$ decomposes into a sum of R simpler functions whose variables separate.

Now, if we discretize the \mathbb{R}^3 space with the above defined subarrays, and take M time samples, we end up with a multi-linear relationship¹ in finite dimensional spaces. In fact, let \mathbf{p}_k be the coordinate vector of the k th sensor of the reference subarray, and $\boldsymbol{\delta}_\ell$ the translation defining the location of the ℓ th subarray, $1 < \ell \leq L$. Then signal (2) can be stored in a $K \times L \times M$ three-way array, which follows the model below, possibly corrupted by additive noise:

$$T_{k\ell m} = \sum_{r=1}^R \lambda_r A_{kr} B_{\ell r} S_{mr} \quad (3)$$

where $A_{kr} = \frac{1}{\sqrt{K}} \exp(j \frac{\omega}{C} \mathbf{p}_k^\top \mathbf{d}_r)$, $B_{\ell r} = \frac{1}{\sqrt{L}} \exp(j \frac{\omega}{C} \boldsymbol{\delta}_\ell^\top \mathbf{d}_r)$, $S_{mr} = \varsigma_r(t_m) / \|\varsigma_r\|$, and $\lambda_r = \sqrt{KL} \|\varsigma_r\|$. Note that \mathbf{a}_r , \mathbf{b}_r and \mathbf{s}_r are hence unit L^2 -norm vectors.

Model (3) is related to the Canonical Polyadic decomposition (CP)², which consists of decomposing a tensor \mathbf{T} into a sum of *decomposable* tensors. For the sake of convenience, equation (3) is rewritten in vector form as

$$\mathbf{t} = \sum_{r=1}^R \lambda_r \mathbf{a}_r \boxtimes \mathbf{b}_r \boxtimes \mathbf{s}_r \quad (4)$$

where \boxtimes denotes the Kronecker³ product, and $\mathbf{t} = \text{vec}\{\mathbf{T}\}$ is a column vector of dimension KLM containing the entries of the 3-way array \mathbf{T} .

¹See *e.g.* [14] for a definition of multi-linearity.

²also sometimes called Candecomp/Parafac in Psychometry or Chemometrics [15].

³This notation is chosen to make the distinction between tensor and Kronecker products, which are sometimes mixed up [14].

III. EXISTENCE AND UNIQUENESS

The goal of this section is to identify the directions of arrival (DoA), ψ_r , of the R impinging plane waves and to estimate corresponding transmitted source signals $\varsigma_r(t_m)$ up to a scaling factor, given the whole array \mathbf{T} . To do this, we shall identify all parameters in the RHS of (4).

A. Low rank approximation

Actually, observations are corrupted by noise, so that (3-4) do not hold exactly. A natural idea is then to fit model (4) by minimizing the error

$$\Upsilon(\mathbf{A}, \mathbf{B}, \mathbf{S}; \boldsymbol{\lambda}) = \left\| \mathbf{t} - \sum_{r=1}^R \lambda_r \mathbf{a}_r \boxtimes \mathbf{b}_r \boxtimes \mathbf{s}_r \right\|^2 \quad (5)$$

where $\boldsymbol{\lambda}$ denotes the vector containing the λ_r 's, and $\|\cdot\|$ the L^2 -norm. Minimizing error (5) means finding the best rank- R approximate of \mathbf{T} and its CP decomposition. However, the infimum of Υ may not be reached; see *e.g.* [16], [11] and references therein. The reason is that the set of rank- R tensors is not closed for $R > 1$. The idea we promote here is to impose an additional constraint that will ensure the existence of a minimum, as elaborated in the next section. In addition, from the physical point of view, one can make the following observations:

- sources that are totally correlated need to be localized separately only if they are sufficiently well angularly separated. In that case they correspond to multi-paths of the same radiating source.
- sources that are located in the same direction need to be estimated separately only if they are sufficiently decorrelated. In the latter case, they correspond to different sources.
- otherwise, one can assimilate highly correlated sources arriving from close directions to a single *fat source*, spread out in space.

The purpose of the section is to formalize these constraints.

B. Coherences

As in the compressed sensing literature [17], [18], we define the *coherence* of a set of unit norm vectors as the largest value of cross scalar products:

$$\mu_A = \sup_{k \neq \ell} |\mathbf{a}_k^H \mathbf{a}_\ell|. \quad (6)$$

The coherence of matrix \mathbf{A} is defined this way, if \mathbf{a}_k denote its (unit norm) columns. Coherences of matrices \mathbf{B} and \mathbf{S} are defined similarly, and denoted by μ_B and μ_S , respectively. Let \mathbf{G} be the $R \times R$ Gram matrix defined by:

$$G_{pq} = (\mathbf{a}_p^H \mathbf{a}_q)(\mathbf{b}_p^H \mathbf{b}_q)(\mathbf{s}_p^H \mathbf{s}_q).$$

Then for given matrices \mathbf{A} , \mathbf{B} and \mathbf{S} , the optimal value $\boldsymbol{\lambda}^o$ minimizing error Υ is obtained by cancelling the gradient of (5) w.r.t. $\boldsymbol{\lambda}$, which leads to the linear system:

$$\mathbf{G} \boldsymbol{\lambda}^o = \mathbf{f}, \quad (7)$$

where vector \mathbf{f} in the right hand side is defined by the contraction $f_r = \sum_{ijk} T_{ijk} A_{ir}^* B_{jr}^* S_{kr}^*$, $1 \leq r \leq R$. Equation (7) shows that coherences play a role in the conditioning of the minimization problem. Also note that only the *product* between coherences appears, and not coherences individually.

C. Existence

We are now in a position to state a sufficient condition of existence of a best rank- R approximation. It has been shown in [11] that if the search set is restricted to

$$\mu_A \mu_B \mu_S \leq \frac{\nu}{R-1} \quad (8)$$

for some $\nu \in (0, 1)$, then the infimum of (5) is reached. This happens because error (5) becomes coercive as soon as (8) is satisfied. And it must then reach its minimum since it is continuous.

Condition (8) can be interpreted as follows [11]. A small value of a coherence, say μ_A , means a sufficient angular separation between columns of matrix \mathbf{A} . For instance, if \mathbf{A} contains steering vectors, then sources are not located in the same physical direction in space; if it contains time signals, then sources are not fully correlated.

Example 1: If $\mu_A = \mu_B = \mu_C = 0.6$, and $R = 3$, then $\mu_A \mu_B \mu_C = 0.216$ is smaller than $1/2$. Condition (8) is hence satisfied for $\nu = 1$. This is the case of Experiment 1, where random matrices are generated with coherences in the interval $[0.2, 0.6]$.

Example 2: If $\mu_A = 0.997$, $\mu_B = 0.999$, $\mu_C = 0.99999$ and $R = 3$, then $\mu_A \mu_B \mu_C = 0.9960$, which is larger than $1/2$. This is the case of Experiment 2.

Remark 1: Condition (8) ensures the existence of a best rank- R approximate to *any* tensor. It is sufficient in the sense that there may exist a tensor which does not satisfy (8) and which still admits a best rank- R approximate.

Remark 2: If the entries of factor matrices are subject to constraints, which is the case in the present context for matrices \mathbf{A} and \mathbf{B} , then the bound (8) becomes loose and could be tightened. In fact, in the proof [11], only straightforward inequalities are used, and in particular the fact that $|\mathbf{a}_p^H \mathbf{a}_q| \leq \mu_A$, which holds by definition of μ_A . In the present framework, columns \mathbf{a}_p belong to the so-called array manifold, which means that there exist some differentiable function $\tilde{\mathbf{a}}(\psi)$ such that $\mathbf{a}_p = \tilde{\mathbf{a}}(\psi_p)$; the exact form of function $\tilde{\mathbf{a}}$ is not important for now, but will be given in (12). Yet, it is clear that $\mu_{\tilde{\mathbf{A}}} = \max_{p < q} |\tilde{\mathbf{a}}(\psi_p)^H \tilde{\mathbf{a}}(\psi_q)| \leq \max_{p < q} |\mathbf{a}_p^H \mathbf{a}_q| = \mu_A$. The same reasoning holding true for matrix \mathbf{B} , one can observe that the existence condition (8) could be replaced by a less constraining one:

$$\mu_{\tilde{\mathbf{A}}} \mu_{\tilde{\mathbf{B}}} \mu_S \leq \frac{\nu}{R-1}$$

where $\mu_{\tilde{\mathbf{A}}} \leq \mu_A$ and $\mu_{\tilde{\mathbf{B}}} \leq \mu_B$. For the sake of simplicity, this will not be subsequently exploited.

Constraint (8) needs some care because it involves max operators, which are not differentiable. For this reason, we propose to use the fact that the L^∞ norm can be bounded by $L^{2\rho}$ norms, and approximated for large values of ρ :

$$\|\mathbf{z}\|_\infty = \max_k \{z_k\} \leq \|\mathbf{z}\|_{2\rho} = \left(\sum_k z_k^{2\rho} \right)^{1/2\rho}, \quad \forall \rho \geq 1$$

for $z_k \in \mathbb{R}^+$. Applying this inequality to $z_k \equiv |\mathbf{a}_p^H \mathbf{a}_q|$ allows to bound coherences above by a differentiable quantity, so that another (somewhat more constraining) sufficient condition can be obtained. More precisely:

$$\mu_A \leq \mu(\mathbf{A}, \rho) \stackrel{\text{def}}{=} \left(\sum_{p < q} |\mathbf{a}_p^H \mathbf{a}_q|^{2\rho} \right)^{1/2\rho}$$

We subsequently call \mathcal{C}_ρ the constraint obtained by replacing the max operators by the $L^{2\rho}$ norms in constraint (8):

$$\mathcal{C}_\rho \stackrel{\text{def}}{=} (1 - R)/\nu + \mu(\mathbf{A}, \rho)^{-1} \mu(\mathbf{B}, \rho)^{-1} \mu(\mathbf{S}, \rho)^{-1} \geq 0 \quad (9)$$

It is clear that if (9) is satisfied, then so is (8). The above is thus a sufficient condition.

D. Uniqueness

There exist sufficient conditions ensuring that the solution of (4) is unique, which involve coherences [11]. However, the condition below is much less constraining [19, p.13]:

$$R \leq M \text{ and } R(R - 1) \leq \frac{K(K - 1)L(L - 1)}{2} \quad (10)$$

and guarantees that there exists almost surely a unique solution. Other sufficient conditions for generic uniqueness exist [20], [21], but may be less attractive when one dimension (*i.e.* M) is large.

IV. JOINT DOA AND SOURCE ESTIMATION

A. Optimization

The constrained optimization is carried out with the help of gradient descent type algorithms, which handle constraints in different manners. Denote for conciseness $\mathbf{x} = \text{vec}\{[\mathbf{A}^T, \mathbf{B}^T, \mathbf{S}^T]\}$ and define the objective function:

$$\mathcal{F}_\rho(\mathbf{x}; \boldsymbol{\lambda}) = \Upsilon(\mathbf{x}; \boldsymbol{\lambda}) + \eta \exp(-\gamma \mathcal{C}_\rho(\mathbf{x})) \quad (11)$$

where η is the penalty weight, γ is introduced to control the sharpness of penalty $\mathcal{C}_\rho(\mathbf{x})$, and $\boldsymbol{\lambda}$ is defined in (7) and depends on \mathbf{x} and \mathbf{t} . This leads to the algorithm below

ALGORITHM

- 1) Choose R satisfying (10).
- 2) Initialize $(\mathbf{A}(0), \mathbf{B}(0), \mathbf{S}(0))$ to matrices with unit-norm columns satisfying $\mathcal{C}_\rho > 0$.
- 3) Compute $\mathbf{G}(0)$ and $\mathbf{f}(0)$, and solve $\mathbf{G}(0) \boldsymbol{\lambda}(0) = \mathbf{f}(0)$ for $\boldsymbol{\lambda}$, according to (7)
- 4) For $k \geq 1$ and subject to a stopping criterion, do
 - a) Compute the descent direction as the gradient w.r.t. \mathbf{x} :
$$\mathbf{d}(k) = -\nabla \mathcal{F}_\rho(\mathbf{x}(k-1); \boldsymbol{\lambda}(k-1))$$
 - b) Compute a stepsize $\ell(k)$
 - c) Update $\mathbf{x}(k) = \mathbf{x}(k-1) + \ell(k) \mathbf{d}(k)$
 - d) Extract the 3 blocks from $\mathbf{x}(k)$: $\mathbf{A}(k)$, $\mathbf{B}(k)$ and $\mathbf{S}(k)$

- e) Normalize the columns of $\mathbf{A}(k)$, $\mathbf{B}(k)$ and $\mathbf{S}(k)$
- f) Compute $\mathbf{G}(k)$ and $\mathbf{f}(k)$, and solve $\mathbf{G}(k) \boldsymbol{\lambda}(k) = \mathbf{f}(k)$ for $\boldsymbol{\lambda}$, according to (7).
- 5) Output $(\mathbf{A}, \mathbf{B}, \mathbf{S}; \boldsymbol{\lambda}) = (\mathbf{A}(k), \mathbf{B}(k), \mathbf{S}(k); \boldsymbol{\lambda}(k))$
- 6) Compute the DoAs from \mathbf{A} as detailed in Section IV-B

In the algorithm, η is decreased as the reconstruction error $\Upsilon(\mathbf{x}; \boldsymbol{\lambda})$ decreases, whereas γ is kept fixed. The step size is determined in a standard way, namely following Wolfe conditions; see *e.g.* [22] for more details.

We give now some gradient expressions⁴ necessary to determine the descent direction $\mathbf{d}(k)$ when \mathcal{F}_ρ is used:

$$\begin{aligned} \frac{\partial \Upsilon}{\partial \mathbf{A}} &= 2 \mathbf{A} \mathbf{M}^A - 2 \mathbf{N}^A \\ \text{with } M_{pq}^A &\stackrel{\text{def}}{=} \sum_{jk} \lambda_p B_{jp} S_{kp}^* B_{jq}^* \lambda_q^* \\ N_{ip}^A &\stackrel{\text{def}}{=} \sum_{jk} T_{ijk} B_{jp}^* S_{kp}^* \lambda_p^* \end{aligned}$$

and

$$\frac{\partial \exp(-\gamma \mathcal{C}_\rho)}{\partial \mathbf{A}} = \frac{\gamma}{\exp(\gamma \mathcal{C}_\rho)} \mathcal{L}_\rho^A \mathbf{A} [(\mathbf{A}^H \mathbf{A}) \boxtimes \boldsymbol{\Omega}^A - \mathbf{I}]$$

where \boxtimes denotes the Hadamard entry-wise product,

$$\mathcal{L}_\rho^A \stackrel{\text{def}}{=} \left(\sum_{q < p} |\mathbf{a}_p^H \mathbf{a}_q|^{2\rho} \right)^{\frac{-1}{2\rho} - 1} \mu(\mathbf{B}, \rho)^{-1} \mu(\mathbf{S}, \rho)^{-1},$$

and $\Omega_{pq}^A \stackrel{\text{def}}{=} |\mathbf{a}_q^H \mathbf{a}_p|^{2\rho-2}$. Keep in mind that expressions above hold true because matrix \mathbf{A} has unit-norm columns. And expressions are similar for matrices \mathbf{B} and \mathbf{S} , which also have unit-norm columns.

B. Source localization

Since in our algorithm, matrix \mathbf{A} has been computed without forcing any structure, it is necessary to extract DoAs from the entries of \mathbf{A} . This approach is suboptimal, but allows to considerably reduce the computational burden.

In the present framework, we consider identical subarrays that are formed of sensors with arbitrary geometries. Each subarray is designed in such a way that the distance between two successive sensors (k and $k+1$) is less than or equal to $\lambda/2$. The elements of \mathbf{A} are given by:

$$A_{kr} = \frac{1}{\sqrt{K}} \exp \left\{ j \frac{2\pi}{\lambda} (p_{k,2} \sin \psi_r + p_{k,3} \cos \psi_r) \right\} \quad (12)$$

where $\lambda = 2\pi C/\omega$ is the wavelength and ψ_r is the DoA of the r th source as illustrated in Figure 1. To exploit this property and avoid phase unwrapping implementation, we form the quotient between two successive elements in the r th column of \mathbf{A} :

$$\begin{aligned} \frac{A_{k+1,r}}{A_{k,r}} &= \exp \left\{ j \frac{2\pi}{\lambda} ((p_{k+1,2} - p_{k,2}) \sin \psi_r + \right. \\ &\quad \left. (p_{k+1,3} - p_{k,3}) \cos \psi_r) \right\} \end{aligned} \quad (13)$$

⁴Matrix gradients are written with the conventions described in [23], [24].

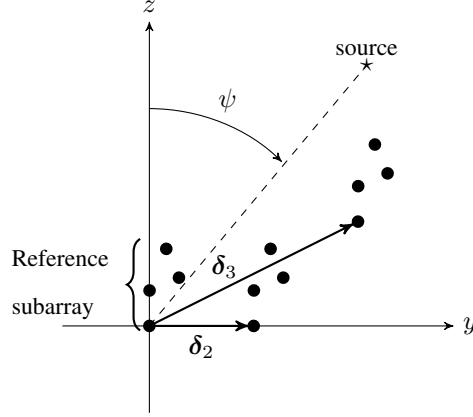


Fig. 1. One source ($R = 1$) radiating on a sensor array with $L = 3$ subarrays.

Then the equation below

$$(p_{k+1,2} - p_{k,2}) \sin \psi_r + (p_{k+1,3} - p_{k,3}) \cos \psi_r = -j \frac{\lambda}{2\pi} \ln \left(\frac{A_{k+1,r}}{A_{k,r}} \right)$$

yields the following linear system for every r , $1 \leq r \leq R$:

$$\mathbf{M} \begin{bmatrix} \sin \psi_r \\ \cos \psi_r \end{bmatrix} = \mathbf{v}_r \quad (14)$$

where $\mathbf{M} = [\bar{\mathbf{p}}_{2:K}^2 - \bar{\mathbf{p}}_{1:K-1}^2, \bar{\mathbf{p}}_{2:K}^3 - \bar{\mathbf{p}}_{1:K-1}^3]$, $\mathbf{v}_r = -j \frac{\lambda}{2\pi} \ln (\mathbf{A}_{2:K,r} \oslash \mathbf{A}_{1:K-1,r})$, symbol \oslash denoting the element-wise division, and $\bar{\mathbf{p}}^i$ denote the columns of the $K \times 3$ matrix containing sensor coordinates:

$$\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_K]^\top = [\bar{\mathbf{p}}^1, \bar{\mathbf{p}}^2, \bar{\mathbf{p}}^3].$$

If the Least Squares (LS) solution of (14) admits a solution whose norm is reasonably close to 1 (we fixed the range to $[0.8, 1.2]$ in the remainder), then the corresponding source DoA estimate $\hat{\psi}_r$ is given by

$$\begin{bmatrix} \sin \hat{\psi}_r \\ \cos \hat{\psi}_r \end{bmatrix} = \frac{\mathbf{w}_r}{\|\mathbf{w}_r\|}, \quad \mathbf{w}_r = \mathbf{M}^\dagger \mathbf{v}_r \quad (15)$$

where † denotes the Moore-Penrose pseudo-inverse. A unique value of ψ_r is then obtained in the interval $[0, 360^\circ]$. On the other hand, if the norm of \mathbf{w} is out of the range $[0.8, 1.2]$, we proceed by direct fit of (12) by exhaustive search for ψ_r . In fact, in the latter case, the fast procedure (15) has revealed to be inaccurate.

C. Source estimation

Unlike source localization, source estimates can be directly obtained from the third factor matrix \mathbf{S} . The complex scaling indeterminacies in matrices \mathbf{A} and \mathbf{B} are fixed thanks to the fact that their first row contain only ones. More precisely, the following transformation

$$(\hat{\mathbf{a}}_r, \hat{\mathbf{b}}_r, \hat{\mathbf{s}}_r; \hat{\lambda}_r) \leftarrow \left(\frac{1}{A_{1r}} \mathbf{a}_r, \frac{1}{B_{1r}} \mathbf{b}_r, \lambda_r A_{1r} B_{1r} \mathbf{s}_r; 1 \right) \quad (16)$$

fixes all scaling indeterminacies. Therefore, source estimates are simply given by $\hat{\mathbf{s}}_r$, $1 \leq r \leq R$.

V. FRAMEWORK OF COMPLEX CRAMÉR-RAO BOUNDS

When parameters are complex, expressions of Cramér-Rao bounds (CRB) depend on the definition of the complex derivative. Since a real function is never holomorphic (unless it is constant) [23], this definition is necessary; this has been overlooked in [25] but clarified in [12]. Originally, the derivative of a real function $\mathbf{h}(\boldsymbol{\theta}) \in \mathbb{R}^p$ with respect to a complex variable $\boldsymbol{\theta} \in \mathbb{C}^n$, $\boldsymbol{\theta} = \boldsymbol{\alpha} + j\boldsymbol{\beta}$, $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^n$, has been defined as the $p \times n$ matrix [23]:

$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \stackrel{\text{def}}{=} \frac{\partial \mathbf{h}}{\partial \boldsymbol{\alpha}} + j \frac{\partial \mathbf{h}}{\partial \boldsymbol{\beta}}$$

Even if the numerical results are independent of the definition assumed for theoretical calculations, we shall subsequently assume the definition proposed in [24], for consistency with [12]:

$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \stackrel{\text{def}}{=} \frac{1}{2} \frac{\partial \mathbf{h}}{\partial \boldsymbol{\alpha}} - \frac{j}{2} \frac{\partial \mathbf{h}}{\partial \boldsymbol{\beta}} \quad (17)$$

With this definition, one has for instance that $\partial \boldsymbol{\alpha} / \partial \boldsymbol{\theta} = \frac{1}{2} \mathbf{I}$, and $\partial \boldsymbol{\beta} / \partial \boldsymbol{\theta} = -\frac{j}{2} \mathbf{I}$ (note the difference with [23], where we had instead $\partial \boldsymbol{\alpha} / \partial \boldsymbol{\theta} = \mathbf{I}$, and $\partial \boldsymbol{\beta} / \partial \boldsymbol{\theta} = j \mathbf{I}$). Assume that parameter $\boldsymbol{\theta}$ is wished to be estimated from an observation \mathbf{t} , of probability distribution $\mathcal{L}(\mathbf{t}; \boldsymbol{\theta})$, and denote $\mathbf{u}(\mathbf{t}; \boldsymbol{\theta})$ the score function. Then we have for any function $\mathbf{h}(\boldsymbol{\theta}) \in \mathbb{R}^p$:

$$\mathbb{E}\{\mathbf{h}(\boldsymbol{\theta}) \mathbf{u}(\mathbf{t}; \boldsymbol{\theta})^\top\} = \frac{\partial}{\partial \boldsymbol{\theta}} \mathbb{E}\{\mathbf{h}(\boldsymbol{\theta})\} \quad (18)$$

with

$$\mathbf{u}(\mathbf{t}; \boldsymbol{\theta})^\top \stackrel{\text{def}}{=} \frac{\partial}{\partial \boldsymbol{\theta}} \log \mathcal{L}(\mathbf{t}; \boldsymbol{\theta}) \quad (19)$$

This is a direct consequence of the fact that $\mathbb{E}\{\mathbf{u}\} = \mathbf{0}$, valid if derivation with respect to $\boldsymbol{\theta}$ and integration with respect to real and imaginary parts of \mathbf{t} can be permuted. Now let $\hat{\boldsymbol{\theta}}(\mathbf{t})$ be an unbiased estimator of $\boldsymbol{\theta}$. Then, following [23], one can prove that $\mathbb{E}\{\hat{\boldsymbol{\theta}} \mathbf{u}^\top\} = \mathbb{E}\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \mathbf{u}^\top\} = \mathbf{I}$ and $\mathbb{E}\{\hat{\boldsymbol{\theta}} \mathbf{u}^H\} = \mathbf{0}$. Finally, by expanding the covariance matrix of the random vector $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) - \mathbf{F}^{-1} \mathbf{u}^*$, one readily obtains that:

$$\mathbf{V} \geq \mathbf{F}^{-1} \quad (20)$$

with

$$\mathbf{V} \stackrel{\text{def}}{=} \mathbb{E}\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^H\} \text{ and } \mathbf{F} \stackrel{\text{def}}{=} \mathbb{E}\{\mathbf{u}^* \mathbf{u}^\top\} \quad (21)$$

Note that the definition of the Fisher information matrix is the complex conjugate of that of [23], because of a different definition of the complex derivation (and hence a different definition of the complex score function). Only notations differ, and bounds on variances remain eventually the same. Expressions (21) are those reported in [24] and are now widely used in signal processing.

VI. DERIVATION OF CRAMÉR-RAO BOUNDS IN THE PRESENCE OF NON CIRCULAR COMPLEX NOISE

For computational purposes, a scaling vector λ has been defined in (7), and offered conditioning advantages. But to ease the derivation of Cramér-Rao bounds, we shall now pull this scaling vector into factor matrices, and consider the model

$$\mathbf{t} = \sum_{r=1}^R \mathbf{a}_r \boxtimes \mathbf{b}_r \boxtimes \mathbf{s}_r, \quad (22)$$

being understood that there are $(K + L + M - 2)R$ free parameters to estimate, since the first entry of \mathbf{a}_r and \mathbf{b}_r is equal to one. More precisely, model (22) is obtained from (4) by the transformations defined in (16).

A. Likelihood

The noise $\mathbf{n} = \mathbf{n}_x + j\mathbf{n}_y$ is assumed to follow a complex normal distribution with zero mean. The likelihood function of \mathbf{t} , defined in (4), takes the form:

$$\begin{aligned} \mathcal{L}(\mathbf{t}, \mathbf{t}^*) &\stackrel{\text{def}}{=} \pi^{-KLM} (\det(\mathbf{\Sigma}) \det(\mathbf{Q}))^{-1/2} \\ &\cdot \exp \left\{ -(\mathbf{t} - \boldsymbol{\mu})^H (\mathbf{Q}^{-1})^* (\mathbf{t} - \boldsymbol{\mu}) + \right. \\ &\quad \left. \text{Re} \left((\mathbf{t} - \boldsymbol{\mu})^T \mathbf{R}^T (\mathbf{Q}^{-1})^* (\mathbf{t} - \boldsymbol{\mu}) \right) \right\} \end{aligned} \quad (23)$$

where $\boldsymbol{\mu}$ is the noise free part of \mathbf{t} , $\mathbf{\Sigma}$ is the covariance matrix of \mathbf{t} , $\mathbf{C} = \mathbb{E} \{ (\mathbf{t} - \boldsymbol{\mu})(\mathbf{t} - \boldsymbol{\mu})^T \}$ is the noncircular covariance (sometimes called *relation matrix*), and

$$\mathbf{Q} = \mathbf{\Sigma}^* - \mathbf{C}^H \mathbf{\Sigma}^{-1} \mathbf{C}, \quad \mathbf{R} = \mathbf{C}^H \mathbf{\Sigma}^{-1} \quad (24)$$

In our computer experiments, a general noise form will be considered, in which non-circularity will be controlled using a variable ε :

$$\text{cov} \{ [\mathbf{n}_x^T, \mathbf{n}_y^T]^T \} = \begin{bmatrix} \mathbf{\Sigma}_{xx} & \mathbf{\Sigma}_{xy} \\ \mathbf{\Sigma}_{yx} & \mathbf{\Sigma}_{yy} \end{bmatrix} = \frac{\sigma^2}{2} \begin{bmatrix} (1 + \varepsilon)\mathbf{I} & \varepsilon\mathbf{I} \\ \varepsilon\mathbf{I} & (1 - \varepsilon)\mathbf{I} \end{bmatrix} \quad (25)$$

In this case $\mathbf{\Sigma} = \sigma^2 \mathbf{I}$ and $\mathbf{C} = \sigma^2 \varepsilon (1 + j)\mathbf{I}$. Therefore, the likelihood function becomes

$$\begin{aligned} \mathcal{L}(\mathbf{t}, \mathbf{t}^*) &= \left(\sigma^2 \pi (1 - 2\varepsilon^2)^{1/2} \right)^{-KLM} \\ &\cdot \exp \left\{ \frac{-1}{\sigma^2 (1 - 2\varepsilon^2)} (\mathbf{t} - \boldsymbol{\mu})^H (\mathbf{t} - \boldsymbol{\mu}) \right. \\ &\quad \left. + \text{Re} \left(\frac{\varepsilon(1 - j)}{\sigma^2 (1 - 2\varepsilon^2)} (\mathbf{t} - \boldsymbol{\mu})^T (\mathbf{t} - \boldsymbol{\mu}) \right) \right\} \end{aligned} \quad (26)$$

Now, let

$$\boldsymbol{\theta} = [\underbrace{\psi_1, \dots, \psi_R}_{\boldsymbol{\psi}}, \underbrace{\bar{\mathbf{b}}_1^T, \dots, \bar{\mathbf{b}}_R^T, \mathbf{s}_1^T, \dots, \mathbf{s}_R^T}_{\boldsymbol{\xi}}, \underbrace{\bar{\mathbf{b}}_1^H, \dots, \mathbf{s}_R^H}_{\boldsymbol{\xi}^*}] \quad (27)$$

denote the unknown parameter vector, where $\bar{\mathbf{b}}_r \stackrel{\text{def}}{=} [B_{2,r}, \dots, B_{L,r}]^T$. Note that, by definition, the likelihood p is a function of both \mathbf{t} and \mathbf{t}^* , and not of $\|\mathbf{t} - \boldsymbol{\mu}\|$ only, especially if noise is non-circularly distributed. In this eventuality, it is hence necessary to introduce both complex parameters and their conjugates in the unknown parameter vector $\boldsymbol{\theta}$, as displayed in (27).

B. Fisher Information

Our goal now is to derive the CRBs of the parameters in $\boldsymbol{\theta}$. The CRBs for factor matrices have been computed in [12]. However, it should be emphasized that, unlike [12], no assumption is needed on the elements of matrix \mathbf{S} to derive the CRB. In fact, assuming that the first row of \mathbf{A} and \mathbf{B} is fixed to $[1, \dots, 1]_{1 \times R}$ is sufficient to reduce the number of parameters to $(K + L + M - 2)R$. Yet, the latter assumption is satisfied in the considered array configuration.

The CRB for unbiased estimation of the complex parameters $\boldsymbol{\theta}$ is equal to the inverse of the Fisher information matrix \mathbf{F} , defined in equation (21). We start by writing the log-likelihood function:

$$\begin{aligned} \log \mathcal{L}(\mathbf{t}, \boldsymbol{\theta}) = & - KLM \log \left(\sigma^2 \pi (1 - 2\varepsilon^2)^{1/2} \right) \\ & - \frac{1}{\sigma^2(1 - 2\varepsilon^2)} (\mathbf{t} - \boldsymbol{\mu})^H (\mathbf{t} - \boldsymbol{\mu}) \\ & + \operatorname{Re} \left(\frac{\varepsilon(1 - j)}{\sigma^2(1 - 2\varepsilon^2)} (\mathbf{t} - \boldsymbol{\mu})^T (\mathbf{t} - \boldsymbol{\mu}) \right) \end{aligned} \quad (28)$$

A straightforward calculation then yields:

$$\begin{aligned} \mathbf{u}^T = & \frac{1}{\sigma^2(1 - 2\varepsilon^2)} \\ & \cdot \left[\mathbf{n}^T \frac{\partial \boldsymbol{\mu}^*}{\partial \boldsymbol{\theta}} + \mathbf{n}^H \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} \right. \\ & \left. - \varepsilon \left((1 + j) \mathbf{n}^H \frac{\partial \boldsymbol{\mu}^*}{\partial \boldsymbol{\theta}} + (1 - j) \mathbf{n}^T \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} \right) \right] \end{aligned} \quad (29)$$

where $\mathbf{n} = \mathbf{t} - \boldsymbol{\mu}$. By substituting the score function \mathbf{u} by its expression, the Fisher information matrix can be written as:

$$\begin{aligned} \mathbf{F} = & \frac{1}{\sigma^2(1 - 2\varepsilon^2)} \\ & \cdot \left[\left(\frac{\partial \boldsymbol{\mu}^*}{\partial \boldsymbol{\theta}} \right)^H \left(\frac{\partial \boldsymbol{\mu}^*}{\partial \boldsymbol{\theta}} - \varepsilon(1 - j) \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} \right) \right. \\ & \left. + \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} \right)^H \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} - \varepsilon(1 + j) \frac{\partial \boldsymbol{\mu}^*}{\partial \boldsymbol{\theta}} \right) \right] \end{aligned} \quad (30)$$

Since parameters in $\boldsymbol{\psi}$ are real and those in $\boldsymbol{\xi}$ are complex, a first writing of the derivatives in (30) is:

$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\psi}}, & \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\xi}}, & \mathbf{0} \end{bmatrix} \quad (31)$$

and

$$\frac{\partial \boldsymbol{\mu}^*}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\psi}} \right)^*, & \mathbf{0}, & \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\xi}} \right)^* \end{bmatrix} \quad (32)$$

Therefore, the Fisher information matrix becomes:

$$\mathbf{F} = \frac{1}{\sigma^2(1 - 2\varepsilon^2)} \begin{bmatrix} 2 \operatorname{Re}\{\mathbf{K}_{11}\} & \mathbf{K}_{12} & \mathbf{K}_{12}^* \\ \mathbf{K}_{12}^H & \mathbf{G}_{22} & \mathbf{H}_{22}^* \\ \mathbf{K}_{12}^T & \mathbf{H}_{22} & \mathbf{G}_{22}^* \end{bmatrix} \quad (33)$$

where

$$\mathbf{K}_{ij} = \mathbf{G}_{ij} + \mathbf{H}_{ij} \quad (34)$$

$$\mathbf{G}_{ij} = \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_i} \right)^H \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_j} \right) \quad (35)$$

$$\mathbf{H}_{ij} = \varepsilon(j-1) \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_i} \right)^T \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_j} \right) \quad (36)$$

with $(i, j) \in \{1, 2\} \times \{1, 2\}$, $\boldsymbol{\theta}_1 = \boldsymbol{\psi}$ and $\boldsymbol{\theta}_2 = \boldsymbol{\xi}$.

To complete the calculation of \mathbf{F} , it remains to give partial derivative expressions of $\boldsymbol{\mu}$ with respect to $\boldsymbol{\psi}$ and $\boldsymbol{\xi}$. This is addressed in the following subsection. To conclude, we present the Fisher information \mathbf{F}_c for a circular complex Gaussian (CCG) noise. In this case, we just need to set $\varepsilon = 0$ in (33), which yields:

$$\mathbf{F}_c = \frac{1}{\sigma^2} \begin{bmatrix} 2 \operatorname{Re}\{\mathbf{G}_{11}\} & \mathbf{G}_{12} & \mathbf{G}_{12}^* \\ \mathbf{G}_{12}^H & \mathbf{G}_{22} & \mathbf{0} \\ \mathbf{G}_{12}^T & \mathbf{0} & \mathbf{G}_{22}^* \end{bmatrix} \quad (37)$$

In view of (37), it is clear that the introduction of $\boldsymbol{\xi}^*$ in the parameter vector is not necessary in the case where the noise follows a CCG distribution. With a non circular complex Gaussian (NCCG) noise, this is not the case.

C. Derivatives of $\boldsymbol{\mu}$ with respect to $\boldsymbol{\psi}$

Using the chain rule we have

$$\frac{\partial \boldsymbol{\mu}}{\partial \psi_f} = \left(\frac{\partial \boldsymbol{\mu}}{\partial \mathbf{a}_f^T} \right) \left(\frac{\partial \mathbf{a}_f^T}{\partial \psi_f} \right) \quad (38)$$

and $[\partial \boldsymbol{\mu} / \partial \mathbf{a}_f^T]$ can be computed using complex derivative formulas. Then, we obtain:

$$\frac{\partial \boldsymbol{\mu}}{\partial \mathbf{a}_f^T} = \mathbf{I}_K \boxtimes \mathbf{b}_f \boxtimes \mathbf{s}_f \in \mathbb{C}^{KLM \times K}, \quad 1 \leq f \leq R. \quad (39)$$

To calculate $[\partial \mathbf{a}_f^T / \partial \psi_f]$, we use the expressions of the considered sensor array configuration, namely equation (12) without the scaling factor $\frac{1}{\sqrt{K}}$, which yields:

$$\frac{\partial \mathbf{a}_f^T}{\partial \psi_f} = -j \frac{2\pi}{\lambda} (\cos \psi_f \bar{\mathbf{p}}^2 - \sin \psi_f \bar{\mathbf{p}}^3) \boxtimes \mathbf{a}_f \quad (40)$$

where $\bar{\mathbf{p}}^2$ and $\bar{\mathbf{p}}^3$ denote the second and third column of the sensor location matrix \mathbf{P} , respectively. By substituting (39) and (40) in (38), we get

$$\begin{aligned} \frac{\partial \boldsymbol{\mu}}{\partial \psi_f} &= -j \frac{2\pi}{\lambda} (\mathbf{I}_K \boxtimes \mathbf{b}_f \boxtimes \mathbf{s}_f) (\cos \psi_f \bar{\mathbf{p}}^2 - \sin \psi_f \bar{\mathbf{p}}^3) \boxtimes \mathbf{a}_f \\ &\stackrel{\text{def}}{=} \boldsymbol{\phi}_{\psi_f} \end{aligned} \quad (41)$$

and

$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\psi}} = [\boldsymbol{\phi}_{\psi_1}, \dots, \boldsymbol{\phi}_{\psi_R}] \in \mathbb{C}^{KLM \times R} \quad (42)$$

D. Derivatives of $\boldsymbol{\mu}$ with respect to $\boldsymbol{\xi}$

Taking partial derivatives of $\boldsymbol{\mu}$ with respect to $\bar{\mathbf{b}}_f^\top$ and \mathbf{s}_f^\top , we obtain:

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \bar{\mathbf{b}}_f^\top} = (\mathbf{a}_f \boxtimes \mathbf{I}_{LM})(\mathbf{I}_L \boxtimes \mathbf{s}_f) \mathbf{J}_L \stackrel{\text{def}}{=} \boldsymbol{\phi}_{\bar{\mathbf{b}}_f} \in \mathbb{C}^{KLM \times (L-1)} \quad (43)$$

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \mathbf{s}_f^\top} = \mathbf{a}_f \boxtimes \mathbf{b}_f \boxtimes \mathbf{I}_M \stackrel{\text{def}}{=} \boldsymbol{\phi}_{\mathbf{s}_f} \in \mathbb{C}^{KLM \times M} \quad (44)$$

where $\mathbf{J}_L = [\mathbf{0}_{(L-1),1} \ \mathbf{I}_{L-1}]^\top \in \mathbb{C}^{L \times (L-1)}$ is a selection matrix. To sum up,

$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\xi}} = [\boldsymbol{\phi}_{\bar{\mathbf{b}}_1}, \dots, \boldsymbol{\phi}_{\bar{\mathbf{b}}_R}, \boldsymbol{\phi}_{\mathbf{s}_1}, \dots, \boldsymbol{\phi}_{\mathbf{s}_R}] \in \mathbb{C}^{KLM \times R(L+M-1)} \quad (45)$$

E. DoA Cramér-Rao bound

The CRB related to DoAs only is obtained as the first leading $R \times R$ block in matrix \mathbf{F}^{-1} , where \mathbf{F} is defined in (33). Doing this assumes that translations $\boldsymbol{\delta}_\ell$ are nuisance parameters, *i.e.* unknown but not of interest. This assumption is relevant in various applications. For instance, consider sonar buoys left floating on the surface and equipped with a device permitting to maintain their orientation towards North. The shape and orientation of subarrays mounted on each buoy are known, but relative locations of buoys are unknown. Similar examples can be encountered when installing arrays of sensors far apart, *e.g.* when performing records on glaciers in difficult conditions: only locations of sensors within subarrays are known accurately. This realistic context has not been considered in the literature. Note that the CRB of the DoA where locations of all sensors are perfectly known can be found in [26], [27].

VII. COMPUTER RESULTS

This section has two goals. First, we show the role of penalty \mathcal{C}_ρ , either along the trajectory when the limiting point lies in the admissible region, or when the constraint is active at the limiting point. This role is emphasized by comparisons with an algorithm of same nature but without penalty. Second, we compare the total Root Mean Square Error (RMSE) to Cramér-Rao bounds. We also evaluate the improvement brought by a third subarray, when comparing our algorithm with ESPRIT (the results obviously coincide with ESPRIT when two subarrays are used).

In all simulations, the sensor array is formed of three shifted identical subarrays of four sensors each, as depicted in Figure 1. The $3 \times R$ matrix defining sensor locations of the reference subarray is

$$\mathbf{P}^\top = \frac{\lambda}{2} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9428 & 0.2814 \\ 0 & 1 & 1.3333 & 2.0833 \end{bmatrix}$$

Next, we set $\rho = 10$, $\gamma = 5$ and $\nu = 1$; these heuristic choices have revealed to be satisfactory in all our experiments.

Remark 3: Theoretically, we should choose $\nu < 1$. But in practice, the choice $\nu = 1$ does not raise any problem, probably for the following reasons: (i) the majoration of the L^∞ norm by the $L^{2\rho}$ norm is strict for almost all values of the argument, so that the bound is loose, (ii) condition (8) is only sufficient, which also leaves some slack, and (iii) even more slack is added by neglecting the fact that \mathbf{a}_r and \mathbf{b}_r are supposed to lie on a manifold, as pointed out in Remark 2.

A. Comparison with other optimization algorithms

To see the interest of the proposed algorithm, we compare it to two other CPD algorithms: i) its homologue without constraint and without computing the optimal value of λ defined in (7), ii) the Levenberg-Marquardt (NLS-LM) gradient decent algorithm [28], [29]. Two performances indices are used: the error on estimated factors and the best sum congruence [30]; this performance index is nothing else but the cosine of the first principal angle [31]. It is defined between two rank-one tensors, $\mathbf{T} = \mathbf{a} \otimes \mathbf{b} \otimes \mathbf{s}$ and $\hat{\mathbf{T}} = \hat{\mathbf{a}} \otimes \hat{\mathbf{b}} \otimes \hat{\mathbf{s}}$, as:

$$\frac{|\langle \mathbf{T}, \hat{\mathbf{T}} \rangle|}{\|\mathbf{T}\| \|\hat{\mathbf{T}}\|} = \frac{|\mathbf{a}^H \hat{\mathbf{a}}|}{\|\mathbf{a}\| \|\hat{\mathbf{a}}\|} \frac{|\mathbf{b}^H \hat{\mathbf{b}}|}{\|\mathbf{b}\| \|\hat{\mathbf{b}}\|} \frac{|\mathbf{s}^H \hat{\mathbf{s}}|}{\|\mathbf{s}\| \|\hat{\mathbf{s}}\|} \quad (46)$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product associated with the Frobenius norm.

To compare two tensors of rank $R > 1$, the *best sum congruence* requires to find the best permutation σ among columns of factor matrices:

$$\max_{\sigma} \sum_{r=1}^R \frac{|\mathbf{a}_r^H \hat{\mathbf{a}}_{\sigma(r)}|}{\|\mathbf{a}_r\| \|\hat{\mathbf{a}}_{\sigma(r)}\|} \frac{|\mathbf{b}_r^H \hat{\mathbf{b}}_{\sigma(r)}|}{\|\mathbf{b}_r\| \|\hat{\mathbf{b}}_{\sigma(r)}\|} \frac{|\mathbf{s}_r^H \hat{\mathbf{s}}_{\sigma(r)}|}{\|\mathbf{s}_r\| \|\hat{\mathbf{s}}_{\sigma(r)}\|} \quad (47)$$

They are compared in two different cases: 1) uncorrelated Gaussian sources with well separated DoAs, *i.e.* the constraint is not active ($\mathcal{C}_\rho \geq 0$) at convergence, and 2) correlated sources with close DoAs, *i.e.* the constraint is active ($\mathcal{C}_\rho < 0$) at convergence. To produce comparable results, the three methods are initialized with the same initial points, and share the same stopping criteria. The latter are:

- the relative change in the objective function. The algorithm stops when

$$\frac{|\mathcal{F}_{\text{current}} - \mathcal{F}_{\text{previous}}|}{\mathcal{F}_{\text{previous}}} \leq 10^{-6}$$

- the tolerance on the Frobenius norm of the gradient divided by the number of entries in the gradient is set to 10^{-8} ,
- the maximum number of iteration is set to 10^3 .

Experiment 1: In this experiment, 1000 independent trials are run with three random well separated DoAs in the interval $[0^\circ, 360^\circ)$, and six randomly generated samples are simulated. Tensors are then of size $4 \times 3 \times 6$. Initial points for the three methods (proposed, unconstrained and NLS-LM) are generated using HOSVD followed by five iterations of ALS. As mentioned in Section IV, η is varied during iterations. More precisely in this experience, η is initialized to 0.1, and is divided by 100 when $\Upsilon(\mathbf{x}; \boldsymbol{\lambda})$ decreases less than 10^{-4} . If the best sum congruence (47) is above a threshold of 0.97 for all R components, then the estimation is considered as correct [32].

Figure 2 reports the number of iterations executed in each trial, where coherences always belonged to the interval $[.2, .6]$, i.e. $\mathcal{C}_\rho > 0$. It can hence be observed that the proposed algorithm converges faster than the unconstrained one. A deeper inspection also reveals that the proposed method yields 99% of correct estimations whereas the unconstrained algorithm and NLS-LM yield 97% and 97.5% respectively.

Figure 3 sketches the evolution of the coherences and $\mathcal{C}_\rho(\mathbf{x})$ as a function of iterations for a typical example where the unconstrained algorithm and NLS-LM do not deliver correct estimates. The figure shows that: (i) the proposed algorithm gives correct estimates thanks to the constraint \mathcal{C}_ρ , which incites iterates to remain inside (or turn back into) the feasible region (where existence is guaranteed), (ii) in the unconstrained algorithm and NLS-LM, iterates run away from the the feasible region and never turn back (the value of \mathcal{C}_ρ remains negative).

Figure 4 presents a comparison of the two algorithms in terms of accuracy of DoAs estimation. It shows the histogram of $\log_{10} \sqrt{\frac{1}{R} \sum_{i=1}^R (\psi_i - \hat{\psi}_i)^2}$ for each trial. We can observe that the proposed algorithm is more accurate than the two others algorithms.

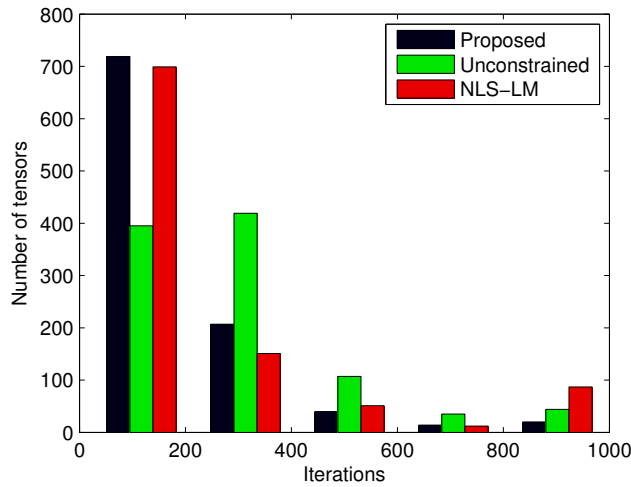


Fig. 2. Number of iterations executed in 1000 independent trials: 99% of the estimations were correct with the proposed algorithm, 97% with the unconstrained algorithm, and 97.5% with the NLS-LM algorithm.

Experiment 2: This experiment intends to show the efficiency of the proposed algorithm when the constraint is active at the limiting point: sources s_2 and s_3 are highly correlated, and closely located. More precisely, we still work with the same array, but only $M = 6$ time samples are used. The tensor size is then $4 \times 3 \times 6$, and

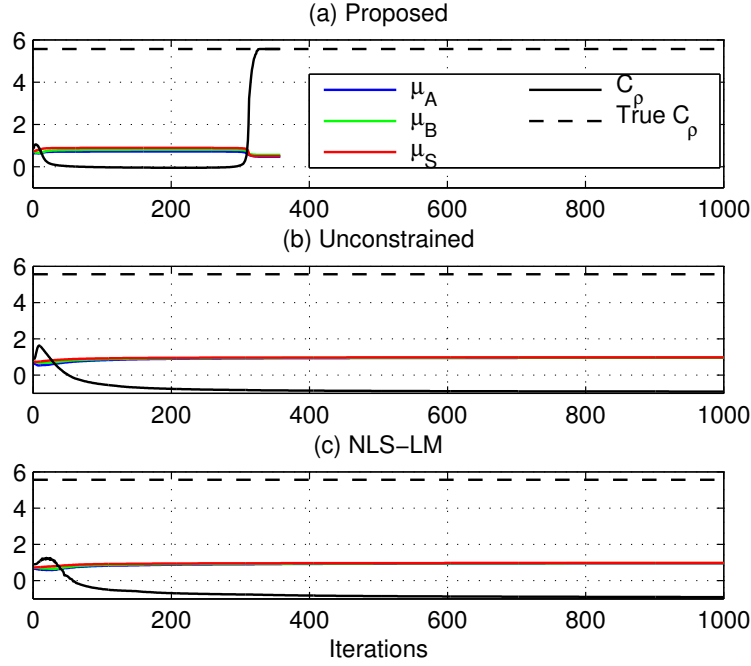


Fig. 3. Coherences (μ_A, μ_B, μ_S) and $C_p(\mathbf{x})$ as a function of the number of iterations. This is a typical example among the trials where the unconstrained optimization and the NLS-LM algorithm diverge.

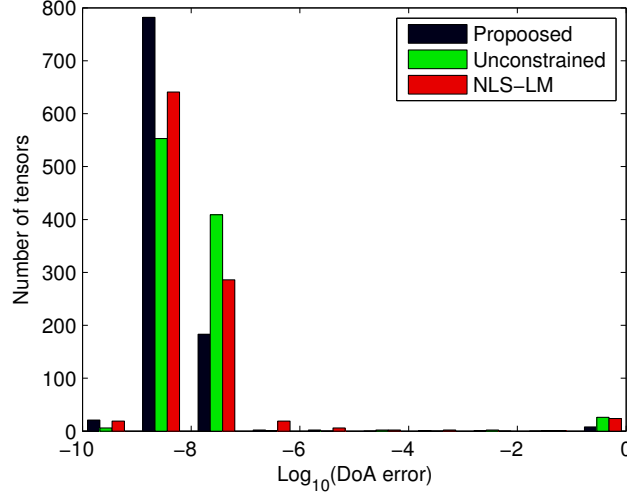


Fig. 4. Estimation error on DoAs for 1000 tensors

$R = 3$ sources impinge on the array, located in directions $\psi_1 = 40^\circ$, $\psi_2 = 83^\circ$ and $\psi_3 = 85^\circ$ respectively. The sources have been generated as follows. One realisation of three 6-dimensional independent random vectors with i.i.d. entries has been drawn, yielding vectors \mathbf{c}_1 , \mathbf{c}_2 and \mathbf{c}_3 . Sources are then built as $\mathbf{s}_1 = \mathbf{c}_1$, $\mathbf{s}_2 = \mathbf{c}_2/\epsilon$, and $\mathbf{s}_3 = \mathbf{c}_2/\epsilon + \mathbf{c}_3$, with $\epsilon = 0.005$. Using these settings, the coherences are then equal to $\mu_A = 0.99$, $\mu_B = 0.99$,

$\mu_S \approx 1$ up to a precision of 10^{-5} , and $\mathcal{C}_\rho = -0.99$. In this experiment, η is initialized to 0.1, and is divided by 10 when the objective decreases less than 10^{-4} between two consecutive iterations, with however a minimal floor value of $\eta_{\min} = 10^{-5}$.

Figure 5 reports results obtained in the noiseless case, with 100 different initial points. It can be noticed that with the same initialisations, the proposed algorithm is able to locate approximately the three sources, whereas the two others algorithms do not even detect the source at 40° . The best sum congruence, averaged over the three sources, is plotted in 5(b); this index reveals poor performances of the unconstrained and the NLS-LM algorithms.

Figure 6 shows the robustness of the proposed algorithm in the presence of low noise. One can observe that sources at $\psi_2 = 83^\circ$ and $\psi_3 = 85^\circ$ are well detected, but suffer from a larger directional bias. This bias comes from the constraint itself, imposing a minimal angular separation.

Experiment 3: This experiment is similar to the previous one except for the tensor rank and the number of time samples. The tensor size is then $4 \times 3 \times 10$, and $R = 4$ sources impinge on the array, located in directions $\psi_1 = 40^\circ$, $\psi_2 = 83^\circ$, $\psi_3 = 85^\circ$ and $\psi_4 = 120^\circ$ respectively. One realization of four 10-dimensional independent random vectors with i.i.d. entries has been drawn, yielding vectors \mathbf{c}_1 , \mathbf{c}_2 , \mathbf{c}_3 and \mathbf{c}_4 . Sources are then built as $\mathbf{s}_1 = \mathbf{c}_1$, $\mathbf{s}_2 = \mathbf{c}_2/\epsilon$, $\mathbf{s}_3 = \mathbf{c}_2/\epsilon + \mathbf{c}_3$ and $\mathbf{s}_4 = \mathbf{c}_4$, with $\epsilon = 0.005$. In this experiment, η is initialized to 1, and is divided by 10 when the objective decreases less than 10^{-2} between two consecutive iterations, with however a minimal floor value of $\eta_{\min} = 10^{-5}$. The results of this experiment are reported in Figure 7. It can be observed that, as in experiment 2, the proposed algorithms outperform the two others.

B. Monte Carlo experiments

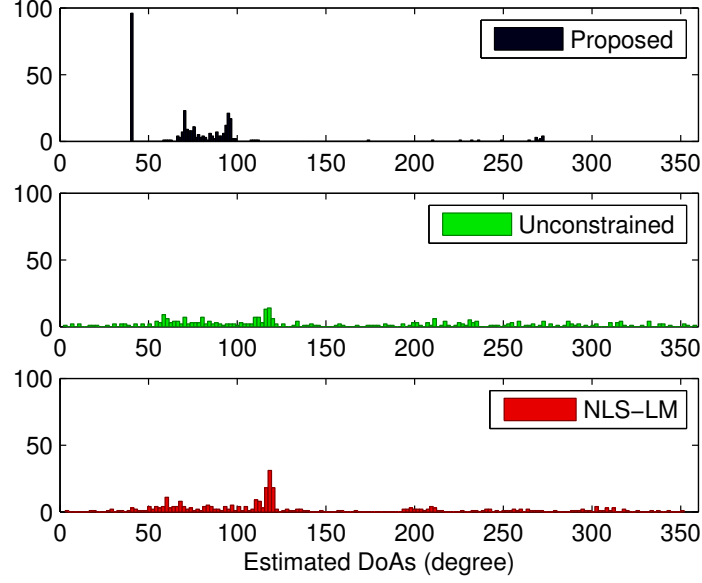
To evaluate the efficiency of the proposed method, we compare its performances to two other algorithms, ESPRIT and MUSIC [26], [27]. The performance criterion is the *total* root mean square error (RMSE) on DoAs defined as:

$$\text{RMSE} = \sqrt{\frac{1}{RN} \sum_{r=1}^R \sum_{n=1}^N (\hat{\psi}_{r,n} - \psi_r)^2}$$

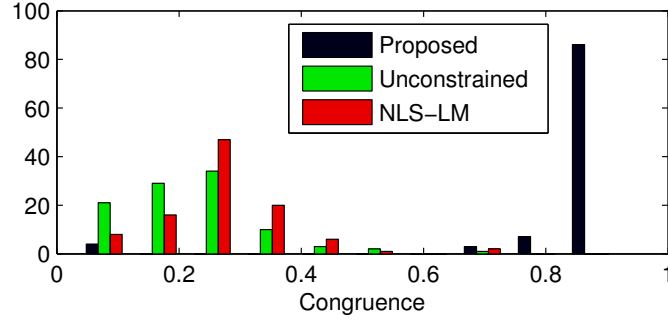
where $\hat{\psi}_{r,n}$ is the estimated DoA in the n -th Monte-Carlo trial and N is the number of trials. The deterministic CRB computed in the previous section is reported as a benchmark. The scenario on which the proposed algorithm is tested can be of interest in numerous applications, where translations δ_ℓ are unknown, as pointed out in Section VI-E. To show the influence of various parameters of the problem on the estimation results, we study three examples, whose parameters are reported in the table below, where $\delta_2 = [0, 25\lambda, 0]^\top$, $\delta_3 = [0, 37.5\lambda, 5\lambda]^\top$.

	Subarrays	Noise	DoA
Exp. 4	$L = 2, (\delta_2)$	CCG	$11^\circ, 57^\circ, 84^\circ$
Exp. 5	$L = 3, (\delta_2, \delta_3)$	CCG	$11^\circ, 57^\circ, 84^\circ$
Exp. 6	$L = 3, (\delta_2, \delta_3)$	NCCG	$11^\circ, 57^\circ, 84^\circ$

In these experiments, the array is still the one depicted in Figure 1, narrowband source signals have the same power, $M = 200$ time samples are used, and 200 Monte-Carlo simulations are run for each SNR level. The value



(a) Histogram of estimated DoAs



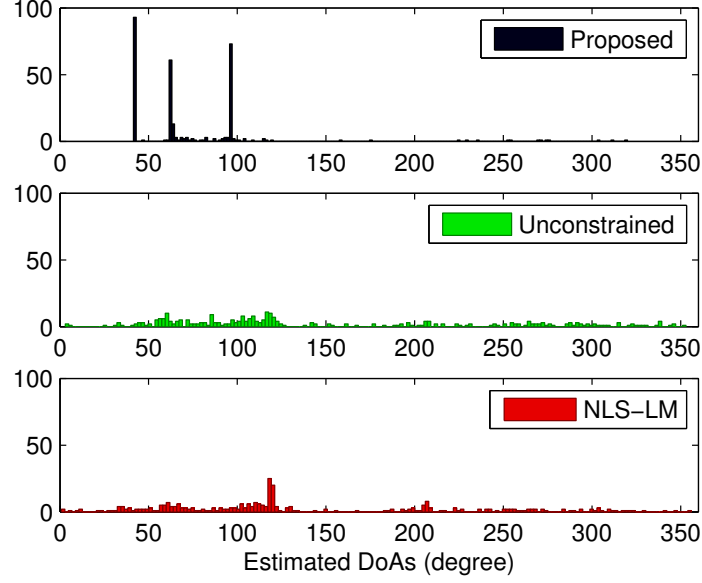
(b) Congruences

Fig. 5. Noiseless scenario with close and correlated sources, $\psi = [40, 83, 85]$. $\mu_A = 0.99$, $\mu_B = 0.99$, $\mu_S \approx 1$ up to a precision of 10^{-5} , and $C_\rho = -0.99$; results with 100 random initializations.

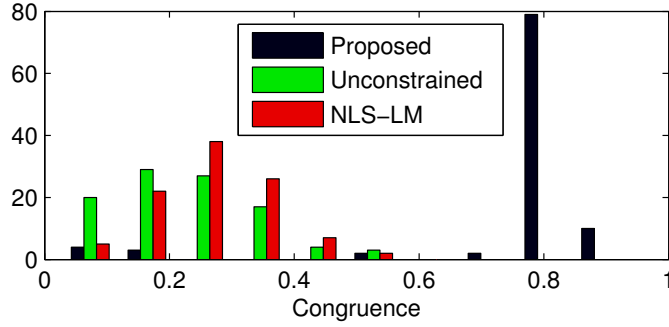
of η is the same as in Exp. 1. Figures 8, 9 and 10 report the RMSE of the DoAs obtained in Exp. 4, 5 and 6, respectively.

Experiment 4: Results reported in Figure 8 show that: (i) the proposed CP algorithm exhibits the same performances as ESPRIT, which makes sense, (ii) MUSIC performs the best, but exploits more information, namely the exact knowledge of sensor locations, whereas this information is actually not available in the present scenario. Hence MUSIC performances just serve as a reference.

Experiment 5: Figure 9 shows that the proposed algorithm yields better results than ESPRIT. The reason is that ESPRIT uses at most two subarrays, whereas the proposed algorithm uses all of them. Again, MUSIC is reported just as a reference benchmark.



(a) Histogram of estimated DoAs

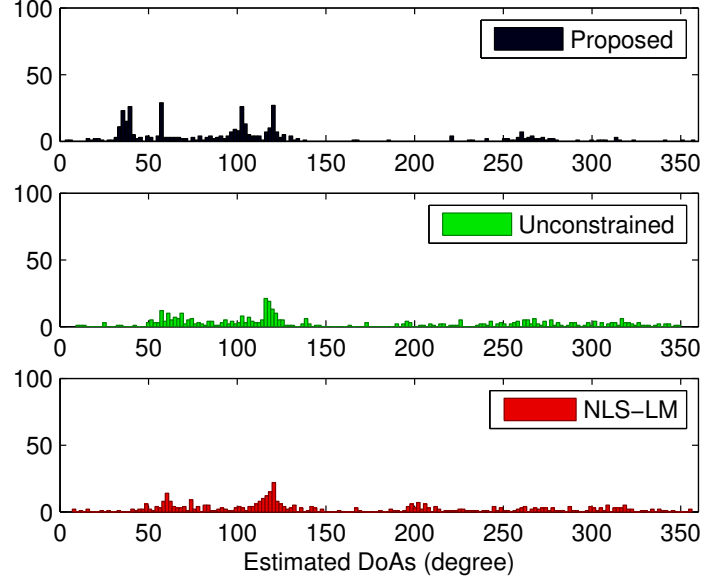


(b) Congruences

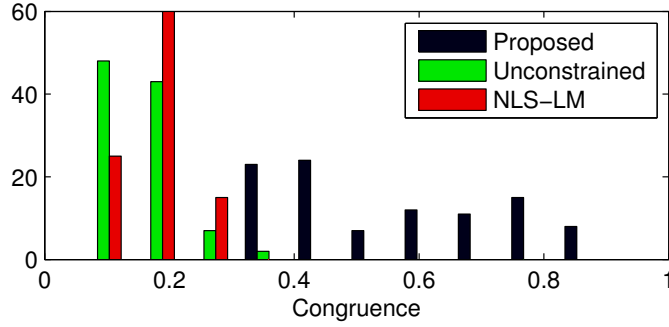
Fig. 6. Noisy scenario with close and correlated sources, $\psi = [40, 83, 85]$. $\mu_A = 0.99, \mu_B = 0.99, \mu_S \approx 1$ up to a precision of 10^{-5} , $C_p = -0.99$ and SNR= 20 dB; results with 100 random initializations.

Experiment 6: Unlike previous experiments, Monte Carlo runs in this example are performed with non-circular complex Gaussian noise with $\varepsilon = 0.1$. The obtained results are similar to those obtained in Experience 5, as depicted in Figure 10.

Experiment 7: To show the influence of the noise distribution on DoA and source estimation, we show results obtained with two values of ε : 0 and 0.5, using the same array configuration and DoA parameters as in Experiment 6. The total RMSE for the parameters estimated by the proposed algorithm is compared to the CRB. Figures 11 and 12 depict, respectively, the results for DoA and source signals, each subfigure corresponding to one value of ε . We conclude that the proposed algorithm yields accurate estimates in the presence of circular as well as noncircular



(a) Histogram of estimated DoAs



(b) Congruences

Fig. 7. Noiseless scenario with close and correlated sources, $\psi = [40, 83, 85, 120]$. $\mu_A = 0.99, \mu_B = 0.99, \mu_S \approx 1$ up to a precision of 10^{-5} , and $C_\rho = -0.99$; results with 100 random initializations.

additive Gaussian noise.

VIII. CONCLUSION

We propose an optimization algorithm including a new differentiable penalty ensuring existence of the low-rank tensor approximation. This general purpose algorithm is applied to antenna array processing as an illustration. DoA estimation of narrow-band far-field sources is formulated as a CP decomposition, when sensor arrays consist of L identical displaced subarrays. We also derive the expressions of the Cramér-Rao bounds of DoA parameters in the presence of nuisance parameters estimated by CP decomposition methods. It is shown that, thanks to our penalty, the proposed algorithm converges quickly and is prevented to leave for long the feasible region. As expected, DoA

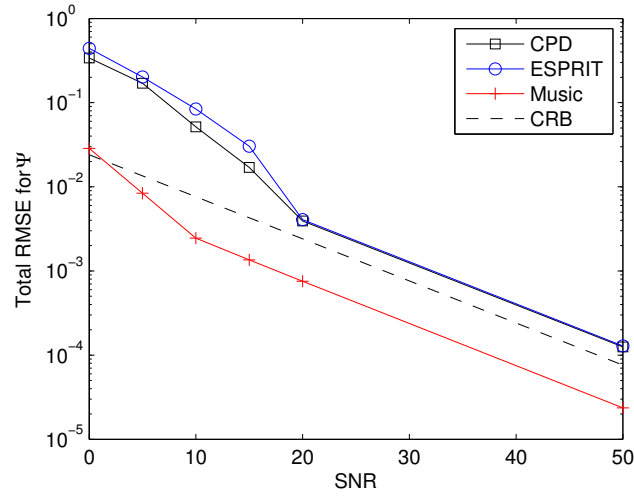


Fig. 8. Total DoA error versus SNR, with $L = 2$ subarrays, $\psi = [11^\circ, 57^\circ, 84^\circ]$.

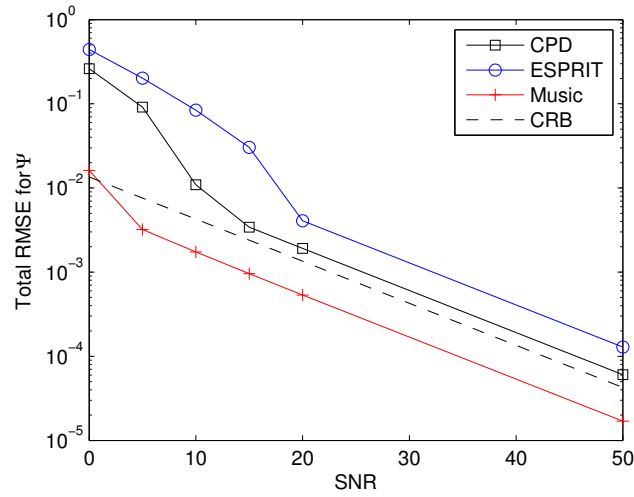


Fig. 9. Total DoA error versus SNR, with $L = 3$ subarrays, $\psi = [11^\circ, 57^\circ, 84^\circ]$.

estimation results show that the CP algorithm exhibits better results than reference DoA estimation methods when $L > 2$.

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for their helpful comments and suggestions.

REFERENCES

- [1] S. Sahnoun and P. Comon, "Deterministic blind identification in antenna array processing," in *8th IEEE SAM Workshop*, A Coruna, Spain, June 22-25 2014, hal-00957357.
- [2] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Sig. Proc. Mag.*, pp. 67–95, July 1996.

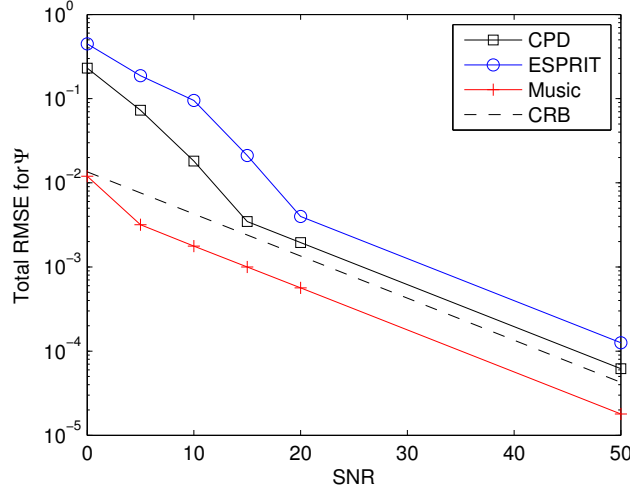


Fig. 10. Total DoA error versus SNR, with $L = 3$ subarrays, $\psi = [11^\circ, 54^\circ, 84^\circ]$, non-circular noise ($\varepsilon = 0.1$).

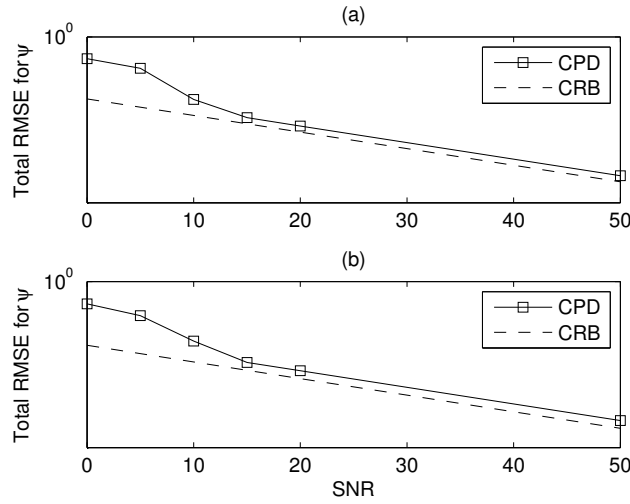


Fig. 11. Total DoA error versus SNR, with $L = 3$ subarrays, $\psi = [11^\circ, 57^\circ, 84^\circ]$. (a) $\varepsilon = 0$, (b) $\varepsilon = 0.5$.

- [3] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629–638, 2011.
- [4] S. Sahnoun, E.-H. Djermoune, C. Soussen, and D. Brie, "Sparse multidimensional modal analysis using a multigrid dictionary refinement," *EURASIP J. Adv. Signal Process.*, March 2012.
- [5] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antenna Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
- [6] H. L. Van Trees, *Optimum Array Processing*, vol. IV, Wiley, New York, 2002.
- [7] R. Roy and T. Kailath, "ESPRIT - estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust. Speech Signal Proc.*, vol. 37, pp. 984–995, July 1989.
- [8] A. L. Swindlehurst, B. Ottersten, R. Roy, and T. Kailath, "Multiple invariance ESPRIT," *IEEE Trans. Signal Proc.*, vol. 40, no. 4, pp. 867–881, Apr. 1992.

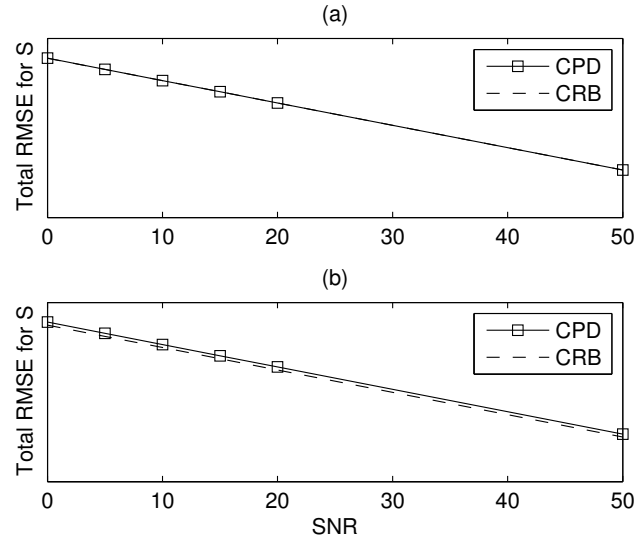


Fig. 12. Total RMSE source error versus SNR, with $L = 3$ subarrays, $\psi = [11^\circ, 57^\circ, 84^\circ]$. (a) $\varepsilon = 0$, (b) $\varepsilon = 0.5$.

- [9] M. Pesavento, A. B. Gershman, and K. M. Wong, "Direction finding in partly calibrated sensor arrays composed of multiple subarrays," *IEEE Trans. Signal Proc.*, vol. 50, no. 9, pp. 2103–2115, Sept. 2002.
- [10] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Sig. Proc.*, vol. 48, no. 8, pp. 2377–2388, Aug. 2000.
- [11] L.-H. Lim and P. Comon, "Blind multilinear identification," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1260–1280, Feb. 2014, open access.
- [12] X. Liu and N. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Trans. Sig. Proc.*, vol. 49, no. 9, pp. 2074–2086, 2001.
- [13] P. Tichavsky, A. H. Phan, and Z. Koldovsky, "Cramér-Rao-induced bounds for Candecomp/Parafac tensor decomposition," *IEEE Trans. on Signal Processing*, 2013.
- [14] P. Comon, "Tensors: a brief introduction," *IEEE Sig. Proc. Magazine*, vol. 31, no. 3, May 2014, special issue on BSS. hal-00923279.
- [15] H. A. L. Kiers, "Towards a standardized notation and terminology in multiway analysis," *J. Chemometrics*, pp. 105–122, 2000.
- [16] P. Comon, "Tensors, usefulness and unexpected properties," in *15th IEEE Workshop on Statistical Signal Processing (SSP'09)*, Cardiff, UK, Aug. 31 - Sep. 3 2009, pp. 781–788, keynote. hal-00417258.
- [17] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Information Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.
- [18] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.
- [19] L. De Lathauwer, "A link between canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," *SIAM Journal on Matrix Analysis*, vol. 28, no. 3, pp. 642–666, 2006.
- [20] M. V. Catalisano, A. V. Geramita, and A. Gimigliano, "Higher secant varieties of the Segre varieties," *Journal of Pure and Applied Algebra*, vol. 201, no. 1-3, pp. 367–380, 2005.
- [21] H. Abo, G. Ottaviani, and C. Peterson, "Induction for secant varieties of Segre varieties," *Trans. Amer. Math. Soc.*, pp. 767–792, 2009, arXiv:math/0607191.
- [22] J. Nocedal and S. J. Wright, *Numerical optimization*, Springer, 2006.
- [23] P. Comon, "Estimation multivariable complexe," *Traitement du Signal*, vol. 3, no. 2, pp. 97–101, Apr. 1986, hal-00979476.

- [24] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Sig. Proc.*, vol. 55, no. 6, pp. 2740–2746, June 2007.
- [25] A. van den Bos, "A Cramér-Rao bound for complex parameters," *IEEE Trans. Sig. Proc.*, vol. 42, no. 10, pp. 2859, Oct. 1994.
- [26] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust. Speech Sig. Proc.*, vol. 37, no. 5, pp. 720–741, 1989.
- [27] B. Ottersten, M. Viberg, and T. Kailath, "Performance analysis of the total least squares ESPRIT algorithm," *IEEE Trans. Sig. Proc.*, vol. 39, no. 5, pp. 1122–1135, 1991.
- [28] K. Madsen, H.B. Nielsen, and O. Tingleff, "Methods for non-linear least squares problems," 2004.
- [29] Pierre Comon, Xavier Luciani, and André LF De Almeida, "Tensor decompositions, alternating least squares and other tales," *Journal of Chemometrics*, vol. 23, no. 7-8, pp. 393–405, 2009.
- [30] G. Tomasi and R. Bro, "A comparison of algorithms for fitting the Parafac model," *Comp. Stat. Data Ana.*, vol. 50, no. 7, pp. 1700–1734, 2006.
- [31] P. Comon and G. H. Golub, "Tracking of a few extreme singular values and vectors in signal processing," *Proceedings of the IEEE*, vol. 78, no. 8, pp. 1327–1343, Aug. 1990.
- [32] A. Smilde, R. Bro, and P. Geladi, *Multi-Way Analysis*, Wiley, Chichester UK, 2004.



Souleymen Sahnoun received the Electrical Engineering degree from the National Polytechnic School of Algiers, Algeria, in 2006, the M.Sc degree from Paul Verlaine University – Metz, France, in 2009 and the Ph.D. degree from the University of Lorraine in 2012.

He is currently a postdoctoral researcher at Gipsa-Lab, affiliated both to the University of Grenoble-Alpes and to CNRS, France. His research interests include statistical signal and array processing, parameter estimation and tensor decompositions.



Pierre Comon (M'87 – SM'95 – F'07) graduated in 1982, and received the Doctorate degree in 1985, both from the University of Grenoble, France. He later received the Habilitation to Lead Researches in 1995, from the University of Nice, France. He has been for nearly 13 years in industry, first with Crouzet-Sextant, Valence, France, between 1982 and 1985, and then with Thomson Marconi, Sophia Antipolis, France, between 1988 and 1997. He spent 1987 with the ISL laboratory, Stanford University, CA. He joined in 1997 the Eurecom Institute, Sophia Antipolis, France. He is research director with CNRS since 1998, first at laboratory I3S, Sophia Antipolis, France, until 2012, and then at Gipsa-Lab, Grenoble, France. His research interests include High-Order Statistics (HOS), Blind techniques, Statistical Signal and

Array Processing, Tensor decompositions, Multi-Way Factor Analysis, Data Mining and its applications to biomedical end environment.

Dr. Comon was Associate Editor of the *IEEE Transactions on Signal Processing* from 1995 to 1998, and a member of the French National Committee of Scientific Research from 1995 to 2000. He was the coordinator of the European Basic Research Working Group on HOS, *ATHOS*, from 1992 to 1995. Between 1992 and 1998, he was a member of the Technical and Scientific Council of the Thomson Group. Between 2001 and 2004 he acted as launching Associate Editor with the *IEEE Transactions on Circuits and Systems I*, in the area of Blind Techniques. He has also been a member of the editorial board of the Elsevier journal *Signal Processing* from 2006 to 2011, and member of several IEEE TC. He is presently in the editorial board of *SIAM Journal on Matrix Analysis and Applications*.